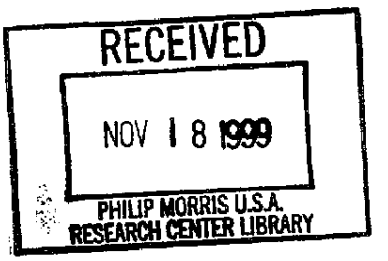
Chapter 1: Basic Concept
Chapter 2: Causation and Causal Inference

Rothman, K. and S. Greenland (eds.) Modern Epidemiology Second Edition
1999, Lippincott-Raven Publishers

PM3003572212

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

RA
651
.R63
1998
c.5



Modern Epidemiology

Second Edition

Kenneth J. Rothman
Professor of Epidemiology
Boston University School of Public Health
Boston University Medical School
Boston, Massachusetts

Sander Greenland
Professor of Epidemiology
Department of Epidemiology
UCLA School of Public Health
Los Angeles, California

with 15 contributors

LIPPINCOTT WILLIAMS & WILKINS
A Wolters Kluwer Company
Philadelphia • Baltimore • New York • London
Buenos Aires • Hong Kong • Sydney • Tokyo

PM3003572213

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

Contents

Contributing Authors ix

Preface to the Second Edition xi

Excerpt from the Preface to the First Edition xii

Acknowledgments xiii

PART I. BASIC CONCEPTS

1. **The Emergence of Modern Epidemiology** 3
Kenneth J. Rothman and Sander Greenland

2. **Causation and Causal Inference** 7
Kenneth J. Rothman and Sander Greenland

3. **Measures of Disease Frequency** 29
Kenneth J. Rothman and Sander Greenland

4. **Measures of Effect and Measures of Association** 47
Sander Greenland and Kenneth J. Rothman

PART II. STUDY DESIGN AND CONDUCT

5. **Types of Epidemiologic Study** 67
Kenneth J. Rothman and Sander Greenland

6. **Cohort Studies** 79
Kenneth J. Rothman and Sander Greenland

7. **Case-Control Studies** 93
Kenneth J. Rothman and Sander Greenland

8. **Precision and Validity in Epidemiologic Studies** 115
Kenneth J. Rothman and Sander Greenland

9. **Accuracy Considerations in Study Design** 135
Kenneth J. Rothman and Sander Greenland

10. **Matching** 147
Kenneth J. Rothman and Sander Greenland

11. **Field Methods in Epidemiology** 163
Patricia Hartge and Jack Cahill

PM3003572214

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

PART III. DATA ANALYSIS

12. **Approaches to Statistical Analysis** 181
Kenneth J. Rothman and Sander Greenland

13. **Fundamentals of Epidemiologic Data Analysis** 201
Sander Greenland and Kenneth J. Rothman

14. **Introduction to Categorical Statistics** 231
Sander Greenland and Kenneth J. Rothman

15. **Introduction to Stratified Analysis** 253
Sander Greenland and Kenneth J. Rothman

16. **Applications of Stratified Analysis Methods** 281
Sander Greenland

17. **Analysis of Polytomous Exposures and Outcomes** 301
Sander Greenland

18. **Concepts of Interaction** 329
Sander Greenland and Kenneth J. Rothman

19. **Basic Methods for Sensitivity Analysis and External Adjustment** 343
Sander Greenland

20. **Introduction to Regression Models** 359
Sander Greenland

21. **Introduction to Regression Modeling** 401
Sander Greenland

PART IV. SPECIAL TOPICS

22. **Surveillance** 435
James W. Buehler

23. **Ecologic Studies** 459
Hal Morgenstern

24. **Analysis of Vital Statistics Data** 481
Suresh H. Moolgavkar, John A. H. Lee, and Richard G. Stevens

25. **Screening** 499
Alan S. Morrison

26. **Clinical Epidemiology** 519
Noel S. Weiss

27. **Concepts of Infectious Disease Epidemiology** 529
M. Elizabeth Halloran

28. **Environmental Epidemiology** 555
Irva Hertz-Picciotto

PM3003572215

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

181
201
231
253
281
301
329
istment 343
359
401
435
459
481
ns
499
519
529
555

29. Reproductive Epidemiology 585
Clarice R. Weinberg and Allen J. Wilcox

30. Genetic Epidemiology 609
Muin J. Khoury

31. Nutritional Epidemiology 623
Walter C. Willett

32. Meta-analysis 643
Sander Greenland

References 675

Subject Index 711

Corrections and Suggestions http://members.aol.com/krothman/modepi.htm

PM3003572216

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

PART I

Basic Concepts

PM3003572217

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

1

The Emergence of Modern Epidemiology

Kenneth J. Rothman and Sander Greenland

Epidemiology is still a young science. Although some excellent epidemiologic studies were conducted before the twentieth century, a systematized body of principles by which to design and judge such studies began to form only after the Second World War. These principles evolved in conjunction with an explosion of epidemiologic activity covering a wide range of health problems.

After the war, the United States initiated many large-scale epidemiologic studies. Several of these studies have had far-reaching influence on health. For example, the community-intervention trials of fluoride supplementation in water that were started during the 1940s have led to widespread primary prevention of dental caries (Ast, 1965). The Framingham Heart Study, initiated in 1949, is the most notable of several long-term follow-up studies of cardiovascular disease that have contributed importantly to understanding the causes of this enormous public-health problem (Dawber et al., 1957; Kannel et al., 1961, 1970; McKee et al., 1971). This remarkable study is continuing to produce valuable findings more than 40 years after it was begun (Kannel and Abbott, 1984; Sytkowski et al., 1990). Knowledge from this and similar epidemiologic studies has helped stem the modern epidemic of cardiovascular mortality in the United States, which peaked in the mid-1960s (Stallones, 1980). The largest formal human experiment ever conducted was the Salk vaccine field trial in 1954, with nearly a million school children as subjects (Francis et al., 1957). This study provided the practical basis for the prevention of paralytic poliomyelitis.

The same era saw the first epidemiologic studies of the impact of tobacco use that eventually led to the landmark report, *Smoking and Health*, issued by the Surgeon General (United States Department of Health, Education and Welfare, 1964). Since that time epidemiologic research has steadily attracted public attention. The news media, boosted by a rising tide of social concern about environmental issues and health in general, have vaulted many epidemiologic studies to prominence. Some of these studies were controversial, although in many cases the media may have been partly responsible for fueling the controversy. A few of the biggest attention-getters were studies related to

- the efficacy of oral antidiabetic medication
- the effect of diethylstilbestrol (DES) on offspring
- clustering and infectious transmission of Hodgkin's disease
- reserpine and breast cancer
- Legionnaires' disease
- low-level ionizing radiation and leukemia
- saccharin and bladder cancer
- swine flu vaccination and Guillain-Barré syndrome

PM3003572218

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

- hormonal drugs in pregnancy and birth defects
- tampons and toxic-shock syndrome
- Bendectin and birth defects
- hazardous waste disposal sites
- replacement estrogens and endometrial cancer
- coffee drinking and pancreatic cancer
- passive smoking
- Agent Orange
- acquired immune deficiency syndrome (AIDS)

Despite the surge of epidemiologic activity in recent years, the evidence indicates that epidemiology remains in an early stage of development. It is only in recent years that epidemiologists have begun to employ consistent definitions for their most basic concepts. In 1975, a paper appeared in the *American Journal of Epidemiology* entitled "Definition of rates: some remarks on their use and misuse" (Elandt-Johnson, 1975). No new concept or definition was proposed, but the paper was useful because so many readers did not know the distinctions among the basic measures used in epidemiology. Clear concepts of causation and related ideas such as induction period are as fundamental to an understanding of epidemiologic research as the definition of basic measures. Nevertheless, even these underpinnings have not yet been fully integrated into the conceptual bedrock of the discipline.

Disagreement about basic conceptual and methodologic points has led in some instances to profound differences in the interpretation of data. In 1978, a controversy erupted about whether exogenous estrogens are carcinogenic to the endometrium: Several case-control studies had reported an extremely strong association, with up to a 15-fold increase in risk, but one group argued that a selection bias accounted for nearly the entire effect (Smith et al., 1975; Ziel and Finkle, 1975; Mack et al., 1976; Horwitz and Feinstein, 1978; Hutchison and Rothman, 1978; Jick et al., 1979; Greenland and Neutra, 1981). Disagreement and confusion about basic ideas in epidemiology do not necessarily attest to the thick-headedness of epidemiologists; a more charitable interpretation would be that the basic ideas fundamental to the new science have not yet displaced traditional thinking.

Why has the explosion in epidemiologic research occurred only recently? The answer lies partially in the difficulty of conducting epidemiologic research. The basic building blocks for epidemiologic inferences are incidence rates. These measures involve counting disease occurrence in relation to the people and time spans in which they occur. They are not easy to obtain. Most diseases occur rarely in human populations, so that considerable time and effort are needed to make the basic measurements.

Epidemiologists also commonly face the problem of obtaining cooperation from numerous other people to make their observations. Unlike experimental science, the investigator cannot manipulate study variables and usually must reckon with limitations imposed by budget and concerns for the privacy of subjects. The end product of such an excruciating and often frustrating exercise is just the first step in accumulating epidemiologic knowledge.

Such difficulties have long discouraged epidemiologic research and will continue to do so. Economies of scale resulting from these observational problems have favored epidemiologic research in settings where medical records and vital statistics are carefully collected and available for use or where the wealth of society can support the expensive efforts needed to gather the necessary information. The logistic problems encountered in

PM3003572219

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

measuring disease incidence have also led to the ascendance of the case-control study as a central tool of modern epidemiology.

Case-control research is in many ways emblematic of the modern synthesis of epidemiologic concepts. The methodology of case-control studies has a sound theoretical basis, and as a means of increasing measurement efficiency in epidemiology, it is an attractive option. Unfortunately, the case-control approach has often been misunderstood to be a second-rate substitute for follow-up studies. Only through a firm conceptual grounding in epidemiologic principles can the student of epidemiology see that there is no basis for blanket derogation of case-control research. Since this type of conceptual grounding, covering a wide range of methodologic issues, is critical to the successful conduct and interpretation of epidemiologic research of all types, a focus on epidemiologic concepts and methods is crucial to anyone who aspires to understand modern epidemiology.

The last third of the twentieth century has seen rapid growth in the understanding and synthesis of epidemiologic concepts. The main stimulus for this conceptual growth seems to have been practice: The explosion of epidemiologic activity accentuated the need to improve understanding of the theoretical underpinnings. For example, the signal studies on smoking and lung cancer in the early 1950s were scientifically noteworthy not only for their substantive findings but also because they demonstrated the efficacy and great efficiency of the case-control study (Wynder and Graham, 1950; Doll and Hill, 1952). Likewise, analysis of data from the Framingham Heart Study stimulated the development of the most popular modeling method in epidemiology today, multiple logistic regression (Cornfield, 1962; Truett et al., 1967).

The fundamental concepts of epidemiology do not, however, depend on empirical results. Thus, the capacity to formulate a theory of epidemiologic concepts has been possible for centuries; that it is a twentieth century phenomenon is independent of any recent scientific and technical breakthroughs. Rather, the economic development of the prosperous nations in the twentieth century afforded the luxury of conducting wide-scale epidemiologic research, which in turn motivated the conceptual development that is the scientific "emergence" of epidemiology.

Until the 1970s, virtually all epidemiologists were physicians. Their interest in epidemiology was typically focused on the occurrence patterns of a particular disease. Perhaps because these researchers subordinated an interest in epidemiologic principles to their substantive goals in understanding disease etiology, there was no movement to pursue the development of a theory of epidemiologic investigation.

Historically, physicians have collaborated fruitfully with statisticians who contributed expertise in making observations on large populations as well as in data analysis. Much of the theoretical development of modern epidemiology was contributed by statisticians: A.B. Hill, J. Cornfield, N. Mantel, N. Breslow, and R.L. Prentice are a few of the outstanding contributors. The influence of statistical thinking in epidemiology has not been wholly positive, however. It was natural for some statisticians, bringing their skills to bear on epidemiologic problems, to borrow methods with which they were familiar in other areas of application. These methods often became incorporated into epidemiologic practice, not always with a sound basis.

One example of the negative influence of statistical thinking in epidemiologic practice is the dominance of statistical hypothesis testing in epidemiologic data analysis. The motivation for the development of statistical hypothesis testing was to provide a basis for decision making in agricultural and quality-control experiments. These experiments were designed to answer questions that called for specific actions, so that the results had to be

PM3003572220

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

classified, if possible, into qualitatively discrete categories. Thus arose the practice of declaring associations in data as "statistically significant" or "nonsignificant," using arbitrary criteria that became conventional. The notion of statistical significance has come to pervade epidemiologic thinking, as well as that of other disciplines. Unfortunately, statistical hypothesis testing is a mode of analysis that offers less insight into epidemiologic data than alternative methods that emphasize estimation of interpretable measures.

Another example of the misapplication of statistics in epidemiology has been in the area of multivariate analysis. Statistical methodology in multivariate modeling has often been transferred wholesale to epidemiology without giving sufficient thought to the underlying epidemiologic concepts. Many practices common in multivariate analysis are often inappropriate in an epidemiologic context: The use of correlation coefficients, stepwise algorithms to determine the model, and variance reduction to evaluate the model are all potentially problematic. Multivariate analysis is an important analytic tool for the epidemiologist, but it cannot be used appropriately without first considering the epidemiologic context that should govern its use.

Today, notwithstanding the important contributions to the field by many who consider themselves first as statisticians or physicians, epidemiologists have achieved a separate identity. Being either a physician or a statistician, or even both simultaneously, is neither a necessary nor sufficient qualification for being an epidemiologist. What is necessary is an understanding of the principles of epidemiologic research and the experience to apply them.

PM3003572221

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

2

Causation and Causal Inference

Kenneth J. Rothman and Sander Greenland

A General Model of Causation
Concept of Sufficient Cause and Component Causes • Strength of Effects • Interaction Among Causes • Proportion of Disease Due to Specific Causes • Induction Period • Generality of the Model
Philosophy of Scientific Inference
Inductivism • Refutationism • Consensus • Bayesianism • Impossibility of Scientific Proof
Causal Inference in Epidemiology
Testing Competing Epidemiologic Theories • Causal Criteria

In *The Magic Years,* Selma Fraiberg (1959) characterizes every toddler as a scientist busily fulfilling an earnest mission to develop a logical structure for the strange objects and events that make up the world that he or she inhabits. To survive successfully requires a useful theoretical scheme to relate the myriad events that are encountered. As a youngster, each person develops and tests an inventory of causal explanations that brings meaning to the events that are perceived and ultimately leads to increasing power to control those events.

Parents can attest to the delight that children take in forming causal hypotheses and then meticulously testing them, often through exasperating repetitions that are motivated mainly by the joy of understanding. Once a child reaches a certain age he will, on entering a new room, search for a wall switch to operate the electric light. On finding one, he will switch it on and off repeatedly to test the discovery beyond any reasonable doubt. Experiments such as those designed to examine the effect of gravity on free-falling liquids are usually conducted with careful attention, varying the initial conditions in subtle ways and reducing extraneous influences whenever possible by conducting the experiments safely removed from parental interference. The fruit of such scientific labors is a working knowledge of the essential system of causal relations that enables each of us to navigate our complex world.

A GENERAL MODEL OF CAUSATION

If everyone begins life as a scientist, creating his or her own inventory of causal explanations for the empirical world, everyone also begins life as a pragmatic philosopher, developing a general causal theory that some events or states of nature are causes with

PM3003572222

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

there would be no skeleton on which to hang the substance of the many specific causal theories that one needs to survive. Unfortunately, the concepts of causation that are established early in life are too rudimentary to serve well as the basis for scientific theories. We need to develop a more refined set of concepts that can serve as a common starting point in discussions of causal theories.

Concept of Sufficient Cause and Component Causes

To begin, we need to define *cause*. We can define a cause of a specific disease event as an antecedent event, condition, or characteristic that was necessary for the occurrence of the disease at the moment it occurred, given that other conditions are fixed. In other words, a cause of a disease event is an event, condition, or characteristic that preceded the disease event and without which the disease event either would not have occurred at all or would not have occurred until some later time. With this definition, it may be that no specific event, condition, or characteristic is sufficient by itself to produce disease. This definition then does not define a complete causal mechanism but a component of it.

A common characteristic of the concept of causation that we develop early in life is the assumption of a one-to-one correspondence between the observed cause and effect. Each cause is seen as necessary and sufficient in itself to produce the effect. Thus, the flick of a light switch appears to be the singular cause that makes the lights go on. There are less evident causes, however, that also operate to produce the effect: the need for an unspent bulb in the light fixture, wiring from the switch to the bulb, and voltage to produce a current when the circuit is closed. To achieve the effect of turning on the light, each of these is equally as important as moving the switch, because absence of any of these components of the causal constellation will prevent the effect.

For many people, the roots of early causal thinking persist and become manifest in attempts to find single causes as explanations for observed phenomena. But experience and reflection should easily persuade us that the cause of any effect must consist of a constellation of components that act in concert (Mill, 1862). A "sufficient cause," which means a complete causal mechanism, can be defined as a set of minimal conditions and events that inevitably produce disease; "minimal" implies that all of the conditions or events are necessary. In disease etiology, the completion of a sufficient cause may be considered equivalent to the onset of disease. (Onset here refers to the onset of the earliest stage of the disease process rather than the onset of signs or symptoms.) For biologic effects, most and sometimes all of the components of a sufficient cause are unknown (Rothman, 1976a).

For example, tobacco smoking is a cause of lung cancer, but by itself it is not a sufficient cause. First, the term *smoking* is too imprecise to be used in a causal description. One must specify the type of smoke (e.g., cigarette, cigar, pipe), whether it is filtered or unfiltered, the manner and frequency of inhalation, and the onset and duration of smoking. More important, smoking, even defined explicitly, will not cause cancer in everyone. So who are those who are "susceptible" to the effects of smoking? Or, to put it in other terms, what are the other components of the causal constellation that act with smoking to produce lung cancer?

When causal components remain unknown, one may be inclined to assign an equal risk to all individuals whose status for some components is known and identical. Thus, men who are heavy cigarette smokers are said to have approximately a 10% lifetime risk of developing lung cancer. Some interpret this statement to mean that all men would be subject to a 10% probability of lung cancer if they were to become heavy smokers, as if the

PM3003572223

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

outcome, aside from smoking, were purely a matter of chance. In contrast, we view the assignment of equal risks as reflecting nothing more than assigning to everyone within a specific category, in this case male heavy smokers, the average of the individual risks for people in that category. In the classic view, these risks are either one or zero, according to whether the individual will or will not get lung cancer.

We cannot measure the individual risks, and assigning the average value to everyone in the category reflects nothing more than our ignorance about the determinants of lung cancer that interact with cigarette smoke. It is apparent from epidemiologic data that some people can engage in chain smoking for many decades without developing lung cancer. Others are or will become "primed" by unknown circumstances and need only to add cigarette smoke to the nearly sufficient constellation of causes to initiate lung cancer. In our ignorance of these hidden causal components, the best we can do in assessing risk is to classify people according to measured causal risk indicators and then assign the average risk observed within a class to persons within the class. As knowledge expands, the risk estimates assigned to people will depart from average according to the presence or absence of other factors that affect the risk.

For example, we now know that smokers with substantial asbestos exposure are at higher risk of lung cancer than those who lack asbestos exposure. Consequently, with adequate data, we could assign different risks to heavy smokers based on their asbestos exposure. Within categories of asbestos exposure, the average risks would be assigned to all heavy smokers until other risk factors are identified.

Figure 2-1 provides a schematic diagram of sufficient causes in a hypothetical individual. Each constellation of component causes represented in Fig. 2-1 is minimally sufficient to produce the disease; that is, there is no redundant or extraneous component cause—each one is a necessary part of that specific causal mechanism. A specific component cause may play a role in one, several, or all of the causal mechanisms.

Figure 2-1 does not depict aspects of the causal process such as prevention, sequence or timing of action of the component causes, dose, and other complexities. These aspects of the causal process must be accommodated in the model by an appropriate definition of each causal component. Thus, if the outcome is lung cancer and the factor E represents cigarette smoking, it might be defined more explicitly as smoking at least two packs a day of unfiltered cigarettes for at least 20 years. If the outcome is smallpox, which is completely prevented by immunization, factor U could represent "unimmunized." More generally, preventive effects of a factor C can be represented by placing its complement "no C" within sufficient causes.

Strength of Effects

The causal model exemplified by Fig. 2-1 can facilitate an understanding of some key concepts such as *strength of effect* and *interaction*. As an illustration of strength of effect,

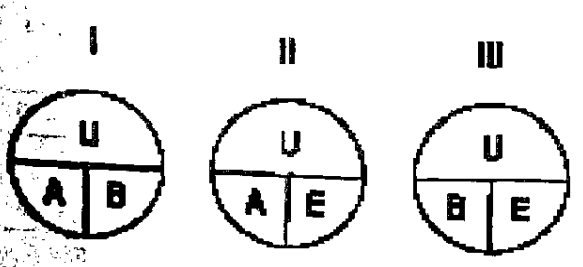


FIG. 2-1. Three sufficient causes of a disease.

PM3003572224

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

TABLE 2-1. *Exposure frequencies for three component causes in two hypothetical populations according to the possible combinations of the component causes*

Exposures			Response (outcome)	Frequency of exposure pattern	
A	B	E		Population 1	Population 2
1	1	1	1	100	900
1	1	0	1	100	900
1	0	1	1	900	100
1	0	0	0	900	100
0	1	1	1	900	100
0	1	0	0	900	100
0	0	1	0	100	900
0	0	0	0	100	900

1, present; 0, absent for exposure and response

Table 2-1 displays the frequency of the eight possible patterns for exposure to A, B, and E in two hypothetical populations. Suppose that U is always present (ubiquitous) and Fig. 2-1 represents all the sufficient causes. Here and throughout the book, we will assume that *disease* refers to a nonrecurrent event, such as death or first occurrence of a disease. Under these assumptions, the response of each individual to the exposure pattern in a given row can be found in the response column.

The proportion getting disease (the incidence proportion) in any subpopulation can be found simply by multiplying the number at each exposure pattern by the response for that pattern, summing these products to get the total number of disease cases in the subpopulation, and dividing this total by the subpopulation size. For example, if exposure A is unmeasured, the pattern of incidence proportions in population 1 would be those in Table 2-2.

As an example of how the proportions in Table 2-2 were calculated, let us review how the incidence proportion among persons with B present but E absent was calculated: There were 100 persons with A present, B present, and E absent, all of whom became cases because A and B are sufficient to produce the disease in combination with the background causes. There were 900 persons with A absent, B present, and E absent, none of whom became cases because they did not have a sufficient cause. Thus, among all 1000 people with B present and E absent, there were 100 cases, for a proportion of 0.10.

It is evident from Table 2-2 that for population 1, E is a much stronger determinant of incidence than B, because the presence of E increases the incidence proportion by 0.9, whereas the presence of B increases it by only 0.1.

Table 2-3 shows the analogous results for population 2. Although the members of this population have exactly the same causal mechanisms operating within them as do the members of population 1, the relative strength of factors E and B are reversed: B is now

TABLE 2-2. *Incidence proportions for combinations of component causes B and E in hypothetical population 1, assuming that component cause A is unmeasured*

	B=1, E=1	B=1, E=0	B=0, E=1	B=0, E=0
Cases	1000	100	900	0
Total	1000	1000	1000	1000
Proportion	1.00	0.10	0.90	0.00

1, present; 0, absent for exposures B and E

PM3003572225

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

TABLE 2-3. *Incidence proportions for combinations of component causes B and E in hypothetical population 2, assuming that component cause A is unmeasured*

	B=1, E=1	B=1, E=0	B=0, E=1	B=0, E=0
Cases	1000	900	100	0
Total	1000	1000	1000	1000
Proportion	1.00	0.90	0.10	0.00

1, present; 0, absent for exposures B and E

a much stronger determinant of incidence than E. This is so despite the fact that in both populations A, B, and E have no association with one another and are each present in exactly half the people.

One key difference between populations 1 and 2 is that the condition under which E acts as a necessary and sufficient cause—the presence of A or B, but not both—is common in population 1 but rare in population 2. In population 1, 3600 people (90% of the total) have A or B but not both, and the incidence proportion for E merely reflects this percentage. In contrast, only 400 people (10% of the total) in population 2 have A or B but not both. This difference in the frequency of the condition necessary and sufficient for E to cause the disease explains the difference in the strength of the effect of E for the two populations. A similar explanation applies to the different strength of effect for factor B in the two populations.

We will call the necessary and sufficient condition for a factor to produce disease the *causal complement* of the factor. Thus, the condition "A or B but not both" is the causal complement of E in the above example. This example shows that the strength of a factor's effect on a population depends on the relative prevalence of its causal complement. This dependence of the effects of a specific component cause on the prevalence of its causal complement has nothing to do with the biologic mechanism of the component's action, since the component is an equal partner in each mechanism in which it appears. Nevertheless, a factor will appear to have a strong effect if its causal complement is common. Conversely, a factor with a rare causal complement will appear to have a weak effect.

In epidemiology, the strength of a factor's effect is usually measured by the change in disease frequency produced by introducing the factor into a population. This change may be measured in absolute or relative terms. In either case, the strength of an effect may have tremendous public-health significance, but it may have little biologic significance. The reason is that, given a specific causal mechanism, any of the component causes can have strong or weak effects. The actual identities of the components of a sufficient cause are part of the biology of causation, whereas the strength of a factor's effect depends on the time-specific distribution of its causal complement in the population. Over a span of time, the strength of the effect of a given factor on the occurrence of a given disease may change because the prevalence of its causal complement in various mechanisms may also change. The causal mechanisms in which the factor and its cofactors act could remain unchanged, however.

Interaction Among Causes

Two component causes acting in the same sufficient cause may be thought of as interacting biologically to produce disease. Indeed, one may define biologic interaction as the

PM3003572226

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

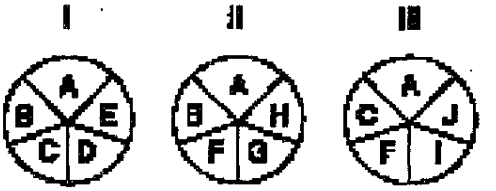


FIG. 2-2. Another example of three sufficient causes of a disease.

participation of two component causes in the same sufficient cause. Such interaction is also known as *causal co-action* or *joint action*. The joint action of the two component causes does not have to be simultaneous action: One component cause could act many years before the other, but it would have to leave some effect that interacts with the later component.

For example, suppose a traumatic injury to the head leads to a permanent disturbance in equilibrium. Many years later, the faulty equilibrium may lead to a fall while walking on an icy path, causing a broken hip. The causal mechanism for the broken hip includes the traumatic injury to the head as a component cause, along with its consequence of a disturbed equilibrium. The causal mechanism also includes the walk along the icy path. These two component causes have interacted with one another, although their time of action is many years apart. They also would interact with the other component causes, such as the type of footwear, the absence of a handhold, and any other conditions that were necessary to the causal mechanism of the fall and the broken hip that resulted.

The degree of observable interaction between two specific component causes depends on how many different sufficient causes produce disease and the proportion of cases that occurs through sufficient causes in which the two component causes both play some role. For example, in Fig. 2-2, suppose that G were only a hypothetical substance that did not actually exist. Consequently, no disease would occur from sufficient cause II because it depends on an action by G, and factors B and F would act only through the distinct mechanisms represented by sufficient causes I and III. Thus, B and F would be biologically independent. Now suppose G is present; then B and F would interact biologically. Furthermore, if C is completely absent, then cases will occur only when factors B and F act together in the mechanism represented by sufficient cause II. Thus, the extent or apparent strength of biologic interaction between two factors is dependent on the prevalence of other factors.

Proportion of Disease Due to Specific Causes

In Fig. 2-1, assuming that the three sufficient causes in the diagram are the only ones operating, what fraction of disease is caused by U? The answer is all of it; without U, there is no disease. U is considered a "necessary cause." What fraction is due to E? E causes disease through two mechanisms, II and III, and all disease arising through either of these two mechanisms is due to E. This is not to say that all disease is due to U alone or that a fraction of disease is due to E alone; no component cause acts alone. Rather, these factors interact with their complementary factors to produce disease.

A widely discussed but unpublished paper from the 1970s written by scientists at the National Institutes of Health proposed that as much as 40% of cancer is attributable to occupational exposures. Many scientists thought that this fraction was unacceptably high and argued against this claim (Higginson, 1980; Ephron, 1984). One of the arguments used in rebuttal was as follows: x% of cancer is caused by smoking, y% by diet, z% by

PM3003572227

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

alcohol, and so on; when all these percentages are added up, only a small percentage, much less than 40%, is left for occupational causes. This rebuttal is fallacious because it is based on the naive view that every case of disease has a single cause. In fact, since diet, smoking, asbestos, and other factors interact with one another and with genetic factors to cause cancer, each case of cancer could be attributed to many separate component causes.

There is a tendency to think that the sum of the fractions of disease attributable to each of the causes of the disease should be 100%. For example, in their widely cited work, *The Causes of Cancer,* Doll and Peto (1981; Table 20) created a table giving their estimates of the fraction of all cancers caused by various agents; the total for the fractions was nearly 100%. Although they acknowledged that any case could be caused by more than one agent (which would mean that the attributable fractions would not sum to 100%), they referred to this situation as a "difficulty" and an "anomaly." It is, however, neither a difficulty nor an anomaly, but simply a consequence of allowing for the fact that no event has a single agent as the cause. The fraction of disease that can be attributed to each of the causes of disease in all the causal mechanisms has no upper limit: For cancer or any disease, the upper limit for the total of the fraction of disease attributable to all the component causes of all the causal mechanisms that produce it is not 100% but infinity. Only the fraction of disease attributable to a single component cause cannot exceed 100%.

A single cause or category of causes that is present in every sufficient cause of disease will have an attributable fraction of 100%. Much publicity attended the pronouncement in 1960 that as much as 90% of cancer is environmentally caused (Higginson, 1960). Since "environment" can be thought of as an all-embracing category that represents nongenetic causes, which must be present to some extent in every sufficient cause, it is clear on *a priori* grounds that 100% of any disease is environmentally caused. Thus, Higginson's estimate of 90% was an underestimate.

Similarly, one can show that 100% of any disease is inherited. MacMahon (1968) cited the example given by Hogben (1933) of yellow shanks, a trait occurring in certain genetic strains of fowl fed on yellow corn. Both the right set of genes and the yellow corn diet are necessary to produce yellow shanks. A farmer with several strains of fowl who feeds them all only yellow corn would consider yellow shanks to be a genetic condition, since only one strain would get yellow shanks, despite all strains getting the same diet. A different farmer who owned only the strain liable to get yellow shanks but who fed some of the birds yellow corn and others white corn would consider yellow shanks to be an environmentally determined condition because it depends on diet. In reality, yellow shanks is determined by both genes and environment; there is no reasonable way to allocate a portion of the causation to either genes or environment. Similarly, every case of every disease has some environmental and some genetic component causes, and therefore every case can be attributed both to genes and to environment. No paradox exists as long as it is understood that the fractions of disease attributable to genes and to environment overlap with one another.

Many researchers have spent considerable effort in developing heritability indices that are supposed to measure the fraction of disease that is inherited. Unfortunately, these indices only assess the relative role of environmental and genetic causes of disease in a particular setting. For example, some genetic causes may be necessary components of every causal mechanism. If everyone in a population has an identical set of the genes that cause disease, however, their effect is not included in heritability indices, despite the fact that having these genes is a cause of the disease. The two farmers in the earlier example would offer very different values for the heritability of yellow shanks, despite the fact that the condition is always 100% dependent on having certain genes.

PM3003572228

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

If all genetic factors that determine disease are taken into account, whether or not they vary within populations, then 100% of disease can be said to be inherited. Analogously, 100% of any disease is environmentally caused, even those diseases that we often consider purely genetic. Phenylketonuria, for example, is considered by many to be purely genetic. Nonetheless, the mental retardation that it may cause can be successfully prevented by appropriate dietary intervention.

The treatment for phenylketonuria illustrates the interaction of genes and environment to cause a disease commonly thought to be purely genetic. What about an apparently purely environmental disease such as "killed in an automobile accident"? It is easy to conceive of genetic traits that lead to psychiatric problems, such as alcoholism, that in turn lead to drunk driving and consequent fatality. Consider another more extreme environmental example, "killed by lightning." Again, partially heritable psychiatric conditions can influence whether someone will take shelter during a lightning storm. The argument may be stretched on this example, but the point that every case of disease has both genetic and environmental causes is theoretically defensible and has important implications for research.

Induction Period

The diagram of causes in Fig. 2-2 also provides a model for conceptualizing the *induction period,* which may be defined as the period of time from causal action until disease initiation. If, in sufficient cause I, the sequence of action of the causes is A, B, C, D, and E and we are studying the effect of B, which (let us assume) acts at a narrowly defined point in time, we do not observe the occurrence of disease immediately after B acts. Disease occurs only after the sequence is completed, so there will be a delay while C, D, and finally E act. When E acts, disease occurs. The interval between the action of B and the disease occurrence is the induction time for the effect of B.

In the example given earlier of an equilibrium disorder leading to a later fall and hip injury, the induction time between the occurrence of the equilibrium disorder and the later hip injury might be very long. In an individual instance, we would not know the exact length of an induction period since we cannot be sure of the causal mechanism that produces disease in an individual instance or when all the relevant component causes acted. We can characterize the induction period relating the action of a component cause to the occurrence of disease in general, however, by accumulating data for many individuals. A clear example of a lengthy induction time is the cause–effect relation between exposure of a female fetus to diethylstilbestrol (DES) and the subsequent development of adenocarcinoma of the vagina. The cancer is usually diagnosed between ages 15 and 30 years. Since the causal exposure to DES occurs early in pregnancy, there is an induction time of about 15–30 years for the carcinogenic action of DES. During this time, other causes presumably are operating; some evidence suggests that hormonal action during adolescence may be part of the mechanism (Rothman, 1981).

It is incorrect to characterize a disease itself as having a lengthy or brief induction time. The induction time can be conceptualized only in relation to a specific component cause. Thus, we say that the induction time relating DES to clear cell carcinoma of the vagina is 15–30 years, but we cannot say that 15–30 years is the induction time for clear cell carcinoma in general. Since each component cause in any causal mechanism can act at a time different from the other component causes, each can have its own induction time. For the component cause that acts last, the induction time equals zero. If another component cause of clear cell carcinoma of the vagina that acts during adolescence were

PM3003572229

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

identified, it would have a much shorter induction time for its carcinogenic action than DES. Thus, induction time characterizes a specific cause–effect pair rather than just the effect.

In carcinogenesis, the terms *initiator* and *promotor* have been used to refer to component causes of cancer that act early and late, respectively, in the causal mechanism. Cancer itself has often been characterized as a disease process with a long induction time. This characterization is a misconception, however, because any late-acting component in the causal process, such as a promotor, will have a short induction time. Indeed, by definition the induction time will always be zero for at least one component cause, the last to act.

Disease, once initiated, will not necessarily be apparent. The time interval between disease occurrence and detection has been termed the *latent period* (Rothman, 1981), although others have used this term interchangeably with induction period. The latent period can be reduced by improved methods of disease detection. The induction period, on the other hand, cannot be reduced by early detection of disease, since disease occurrence marks the end of the induction period. Earlier detection of disease, however, may reduce the apparent induction period (the time between causal action and disease detection), since the time when disease is detected, as a practical matter, is usually used to mark the time of disease occurrence. Thus, diseases such as slow-growing cancers may appear to have long induction periods with respect to many causes because they have long latent periods. The latent period, unlike the induction period, is a characteristic of the disease and the detection effort applied to the person with the disease.

Although it is not possible to reduce the induction period proper by earlier detection of disease, it may be possible to observe intermediate stages of a causal mechanism. The increased interest in biomarkers such as DNA adducts is an example of attempting to focus on causes more proximal to the disease occurrence. Such biomarkers may reflect the effects of earlier-acting agents on the organism.

Some agents may have a causal action by shortening the induction time of other agents. Suppose that exposure to factor A leads to epilepsy after an interval of 10 years, on the average. It may be that exposure to a drug, B, would shorten this interval to 2 years. Is B acting as a catalyst or as a cause of epilepsy? The answer is both: a catalyst is a cause. Without B, the occurrence of epilepsy comes 8 years later than it comes with B, so we can say that B causes the onset of the early epilepsy. It is not sufficient to argue that the epilepsy would have occurred anyway. First, it would not have occurred at that time, and the time of occurrence is part of our definition of an event. Second, epilepsy will occur later only if the individual survives an additional 8 years, which is not certain. Not only does agent B determine when the epilepsy occurs, but it can also determine whether it occurs. Thus, we should call any agent that acts as a catalyst of a causal mechanism, speeding up an induction period for other agents, as a cause in its own right. Similarly, any agent that postpones the onset of an event, drawing out the induction period for another agent, is a preventive. It should not be too surprising to equate postponement to prevention: We routinely use such an equation when we employ the euphemism that we prevent death, which actually can only be postponed. What we prevent is death at a given time, in favor of death at a later time.

Generality of the Model

The main utility of this model of sufficient causes and their components lies in its ability to provide a general but practical conceptual framework for causal problems. The at-

PM3003572230

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

tempt to make the proportion of disease attributable to various component causes add to 100% is an example of a fallacy that is exposed by the model: The model makes it clear that, because of interactions, there is no upper limit to the sum of these proportions. As we shall see in Chapter 18, the epidemiologic evaluation of interactions themselves can be clarified with the help of the model.

How could the model accommodate varying doses of a component cause? Since the model appears to deal qualitatively with the action of component causes, it might seem that dose variability cannot be taken into account. But this view is overly pessimistic. To account for dose variability, one need only postulate a set of sufficient causes, each of which contains as a component a different dose of the agent in question. Small doses might require a larger or rarer set of complementary causes to complete a sufficient cause than that required by large doses (Rothman, 1976a). In this way, the model could account for the phenomenon of a shorter induction period accompanying larger doses of exposure, because there would be a smaller set of complementary components needed to complete the sufficient cause.

Those who believe that chance must play a role in any complex mechanism might object to the intricacy of this deterministic model. A probabilistic (stochastic) model could be invoked to describe a dose–response relation. for example, without the need for a multitude of different causal mechanisms; the model would simply relate the dose of the exposure to the probability of the effect occurring. For those who believe that virtually all events contain some element of chance, deterministic causal models may seem to misrepresent reality. Nevertheless, the deterministic model presented here can accommodate "chance," but it does so by reinterpreting chance as deterministic events beyond the current limits of knowledge or observability.

For example, the outcome of a flip of a coin is usually considered a chance event. In classical mechanics, however, the outcome can in theory be determined completely by the application of physical laws and a sufficient description of the starting conditions. To put it in terms more familiar to epidemiologists, consider the explanation for why an individual gets lung cancer. One hundred years ago. when little was known about the etiology of lung cancer, a scientist might have said that it was a matter of chance. Nowadays, we might say that the risk depends on how much the individual smokes, how much asbestos and radon the individual has been exposed to, and so on. One might then ask what determines whether an individual who has smoked a specific amount and has a specified amount of exposure to all the other known risk factors will get lung cancer. Today's answer might well be that it is a matter of chance. We can explain much more of the variability in lung cancer occurrence nowadays than we formerly could by taking into account factors known to cause it, but at the limits of our knowledge, we ascribe the remaining variability to what we call chance. In this view, chance is seen as a catchall term for our ignorance about causal explanations.

We have so far ignored more subtle considerations of sources of unpredictability in events, such as chaotic behavior (in which even the slightest uncertainty about initial conditions leads to vast uncertainty about outcomes) and quantum-mechanical uncertainty. In each of these situations, a random (stochastic) model component may be essential for any useful modeling effort. Such components can be introduced in the above conceptual model by treating unmeasured component causes in the model as random events, so that the causal model based on components of sufficient causes can have a random element.

PHILOSOPHY OF SCIENTIFIC INFERENCE

Causal inference may be viewed as a special case of the more general process of scientific reasoning. The literature on this topic is too vast for us to review, but we will pro-

PM3003572231

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

vide a brief overview of certain points relevant to epidemiology, at the risk of some oversimplification.

Inductivism

Modern science began to emerge around the sixteenth and seventeenth centuries, when the knowledge demands of emerging technologies (such as artillery and transoceanic navigation) stimulated inquiry into the origins of knowledge. An early codification of the scientific method was Francis Bacon's *Novum Organum*, which, in 1620, presented an inductivist view of science. In this philosophy, scientific reasoning is said to depend on making generalizations, or inductions, from observations to general laws of nature; the observations are said to induce the formulation of a natural law in the mind of the scientist. Thus, an inductivist would have said that Jenner's observation of lack of smallpox among milkmaids induced in Jenner's mind the theory that cowpox (common among milkmaids) conferred immunity to smallpox. Inductivist philosophy reached a pinnacle of sorts in the canons of John Stuart Mill (1862), which evolved into inferential criteria that are still in use today.

Inductivist philosophy was a great step forward from the medieval scholasticism that preceded it, for at least it demanded that a scientist make careful observations of people and nature rather than appeal to faith, ancient texts, or authorities. Nonetheless, by the eighteenth century, the Scottish philosopher David Hume had described a disturbing deficiency in inductivism: An inductive argument carried no logical force; instead, such an argument represented nothing more than an *assumption* that certain events would in the future follow in the same pattern as they had in the past. Thus, to argue that cowpox caused immunity to smallpox because no one got smallpox after having cowpox corresponded to an unjustified assumption that the pattern observed so far (no smallpox after cowpox) will continue into the future. Hume pointed out that, even for the most reasonable-sounding of such assumptions, there was no logic or force of necessity behind the inductive argument.

Of central concern to Hume (1739) was the issue of causal inference and failure of induction to provide a foundation for it:

> Thus not only our reason fails us in the discovery of the *ultimate connexion* of causes and effects, but even after experience has inform'd us of their *constant conjunction*, 'tis impossible for us to satisfy ourselves by our reason, why we shou'd extend that experience beyond those particular instances, which have fallen under our observation. We suppose, but are never able to prove, that there must be a resemblance betwixt those objects, of which we have had experience, and those which lie beyond the reach of our discovery.

In other words, no number of repetitions of a particular sequence of events, such as the appearance of a light after flipping a switch, can establish a causal connection between the action of the switch and the turning on of the light. No matter how many times the light comes on after the switch has been pressed, the possibility of coincidental occurrence cannot be ruled out. Hume pointed out that observers cannot perceive causal connections, but only a series of events. Bertrand Russell (1945) illustrated this point with the example of two accurate clocks that perpetually chime on the hour, with one keeping time slightly ahead of the other; although one invariably chimes before the other, there is no causal connection from one to the other. Thus, assigning a causal interpretation to the pattern of events cannot be a logical extension of our observations, since the events might be occurring together only by coincidence, or because of a shared earlier cause.

Causal inference based on mere coincidence of events constitutes a logical fallacy known as *post hoc ergo propter hoc* (Latin for "after this therefore on account of this").

PM3003572232

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

This fallacy is exemplified by the inference that the crowing of a rooster is necessary for the sun to rise because sunrise is always preceded by the crowing.

The *post hoc* fallacy is a special case of a more general logical fallacy known as the *fallacy of affirming the consequent.* This fallacy of confirmation takes the following general form: "We know that if H is true, B must be true; and we know that B is true; therefore H must be true." This fallacy is used routinely by scientists in interpreting data. It is used, for example, when one argues as follows: "If sewer service causes heart disease, then heart disease rates should be highest where sewer service is available; heart disease rates are indeed highest where sewer service is available; therefore, sewer service causes heart disease." Here, H is the hypothesis "sewer service causes heart disease" and B is the observation "heart disease rates are highest where sewer service is available." The argument is of course logically unsound, as demonstrated by the fact that we can imagine many ways in which the premises could be true but the conclusion false; for example, economic development could lead to both sewer service and elevated heart disease rates, without any effect of sewer service on heart disease.

Bertrand Russell (1945) summarized the fallacy this way:

> 'If *p*, then *q*; now *q* is true; therefore *p* is true.' E.g., 'If pigs have wings, then some winged animals are good to eat; now some winged animals are good to eat; therefore pigs have wings.' This form of inference is called 'scientific method.'

Refutationism

Russell was not alone in his lament of the illogicality of scientific reasoning as ordinarily practiced. Many philosophers and scientists from Hume's time forward attempted to set out a firm logical basis for scientific reasoning. Perhaps none has attracted more attention from epidemiologists than the philosopher Karl Popper.

Popper (1959) addressed Hume's problem by asserting that scientific hypotheses can never be proven or established as true in any logical sense. Instead, Popper observed that scientific statements can simply be found to be consistent with observation. Since it is possible for an observation to be consistent with several hypotheses that themselves may be mutually inconsistent, consistency between an hypothesis and an observation is no proof of the hypothesis. In contrast, a valid observation that is inconsistent with an hypothesis implies that the hypothesis as stated is false, and so refutes the hypothesis. If you wring the rooster's neck before it crows and the sun still rises, you have disproved that the rooster's crowing is a necessary cause of sunrise. Or consider a hypothetical research program to learn the boiling point of water (Magee, 1985). A scientist who boils water in an open flask and repeatedly measures the boiling point at 100°C will never, no matter how many confirmatory repetitions are involved, prove that 100°C is always the boiling point. On the other hand, merely one attempt to boil the water in a closed flask or at high altitude will refute the proposition that water always boils at 100°C.

According to Popper, science advances by a process of elimination that he called "conjecture and refutation." Scientists form hypotheses based on intuition, conjecture, and previous experience. Good scientists use deductive logic to infer predictions from the hypothesis and then compare observations with the predictions. Hypotheses whose predictions agree with observations are confirmed only in the sense that they can continue to be used as explanations of natural phenomena. At any time, however, they may be refuted by further observations and replaced by other hypotheses that better explain the observations. This view of scientific inference is sometimes called *refutationism* or *falsificationism.*

PM3003572233

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

Refutationists consider induction to be a psychologic crutch: Repeated observations did not in fact induce the formulation of a natural law, but only the belief that such a law has been found. For a refutationist, only the psychologic comfort that induction provides explains why it still has its advocates.

One way to rescue the concept of induction from the stigma of pure delusion is to resurrect it as a psychologic phenomenon, as Hume and Popper claimed it was, but one that plays a legitimate role in hypothesis formation. The philosophy of conjecture and refutation places no constraints on the origin of conjectures. Even delusions are permitted as hypotheses, and therefore inductively inspired hypotheses, however psychologic, are valid starting points for scientific evaluation. This concession does not admit a logical role for induction in confirming scientific hypotheses, but it allows the process of induction to play a part, along with imagination, in the scientific cycle of conjecture and refutation.

The philosophy of conjecture and refutation has profound implications for the methodology of science. The popular concept of a scientist doggedly assembling evidence to support a favorite thesis is objectionable from the standpoint of refutationist philosophy because it encourages scientists to consider their own pet theories as their intellectual property, to be confirmed, proven, and, when all the evidence is in, cast in stone and defended as natural law. Such attitudes hinder critical evaluation, interchange, and progress. The approach of conjecture and refutation, in contrast, encourages scientists to consider multiple hypotheses and to seek crucial tests that decide between competing hypotheses by falsifying one of them. Since falsification of one or more theories is the goal, there is incentive to depersonalize the theories. Criticism leveled at a theory need not be seen as criticism of its proposer. It has been suggested that the reason why certain fields of science advance rapidly while others languish is that the rapidly advancing fields are propelled by scientists who are busy constructing and testing competing hypotheses; the other fields, in contrast, "are sick by comparison, because they have forgotten the necessity for alternative hypotheses and disproof" (Platt, 1964).

Consensus

Some twentieth century philosophers of science, most notably Thomas Kuhn (1962), have emphasized the role of the scientific community in determining the validity of scientific theories. These critics of the conjecture and refutation model have suggested that the refutation of a theory involves making a choice. Every observation is itself dependent on theories. For example, observing the moons of Jupiter through a telescope seems to us like a direct observation, but only because the theory of optics on which the telescope is based is so well accepted. When confronted with a refuting observation, a scientist faces the choice of rejecting either the validity of the theory being tested or the validity of the scientific infrastructure of the theories on which the refuting observation is based (or rejecting the refuting observation!). Observations that are falsifying instances of theories may at times be treated as "anomalies," tolerated without falsifying the theory in the hope that the anomalies may eventually be explained. An epidemiologic example is the observation that shallow-inhaling smokers had higher lung cancer rates than deep-inhaling smokers. This anomaly was eventually explained when it was noted that smoking-associated lung tumors tend to occur high in the lung, where shallowly inhaled smoke tars tend to be deposited (Wald, 1985).

In other instances, anomalies may eventually lead to the overthrow of current scientific doctrine, just as Newtonian mechanics was discarded (remaining only as a first-order approximation) in favor of relativity theory. Kuhn claimed that in every branch of science

PM3003572234

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

the prevailing scientific viewpoint, which he termed "normal science," occasionally undergoes major shifts that amount to scientific revolutions. These revolutions signal a decision of the scientific community to discard the scientific infrastructure rather than to falsify a new hypothesis that cannot be easily grafted onto it. Kuhn and others have argued that the consensus of the scientific community determines what is considered accepted and what is considered refuted.

Kuhn's critics characterized this description of science as one of an irrational process, "a matter for mob psychology" (Lakatos, 1970). Those who believe in a rational structure for science consider Kuhn's vision to be a regrettably real description of much of what passes for scientific activity, but not prescriptive for any good science. The philosophic debate about Kuhn's description of science hinges on whether Kuhn meant to describe only what has happened historically in science or instead what ought to happen, an issue about which Kuhn (1970) has not been completely clear:

> Are Kuhn's remarks about scientific development...to be read as descriptions or prescriptions? The answer, of course, is that they should be read in both ways at once. If I have a theory of how and why science works, it must necessarily have implications for the way in which scientists should behave if their enterprise is to flourish.

The idea that science is a sociologic process, whether considered descriptive or normative, is an interesting thesis. Regardless of the answer, we suspect that most epidemiologists (and most scientists) will continue to function as if the following classic view is correct: The ultimate goal of scientific inference is to capture some objective truths about the material world in which we live, and any theory of inference should ideally be evaluated by how well it leads us to these truths.

Those holding the view that scientific truth is not arbitrary nevertheless concede that our knowledge of these truths will always be tentative. For refutationists, this tentativeness has an asymmetric quality: We may know a theory is false because it consistently fails the tests we put it through, but we cannot know that it is true, even if it passes every test we can devise, for it may fail a test as yet undevised. With this view, any theory of inference should ideally be evaluated by how well it leads us to detect errors in our hypotheses and observations.

Bayesianism

There is another philosophy of inference that, like refutationism, holds an objective view of scientific truth and a view of knowledge as tentative or uncertain, but that focuses on evaluation of knowledge rather than truth. Like refutationism, the modern form of this philosophy evolved from the writings of eighteenth century philosophers. The focal arguments first appeared in a pivotal essay by Thomas Bayes (1764), and hence the philosophy is usually referred to as Bayesianism (Howson and Urbach, 1993). Like refutationism, it did not reach a complete expression until after World War I, most notably in the writings of Ramsey (1931) and DeFinetti (1937); and, like refutationism, it did not begin to appear in epidemiology until the 1970s (e.g., Cornfield, 1976).

The central problem addressed by Bayesianism is the following: In classic logic, a deductive argument can provide no information about the truth or falsity of a scientific hypothesis unless you can be 100% certain about the truth of the premises of the argument. Consider the logical argument called modus tollens: "If H implies B, and B is false, then H must be false." This argument is logically valid, but the conclusion follows only on the assumptions that the premises "H implies B" and "B is false" are true statements. If these

PM3003572235

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

premises are statements about the physical world, we cannot possibly know them to be correct with 100% certainty since all observations are subject to error. Furthermore, the claim that "H implies B" will often depend on its own chain of deductions, each with its own premises of which we cannot be certain.

For example, if H is "television viewing causes homicides" and B is "homicide rates are highest where televisions are most common," the first premise used in *modus tollens* to test the hypothesis that television viewing causes homicides will be: "If television viewing causes homicides, homicide rates are highest where televisions are most common." The validity of this premise is doubtful—after all, even if television does cause homicides, homicide rates may be low where televisions are common because of socioeconomic advantages in those areas.

Continuing to reason in this fashion, we could arrive at a more pessimistic state than even Hume imagined: Not only is induction without logical foundation, but *deduction* has no scientific utility because we cannot insure the validity of all the premises. The Bayesian answer to this problem is partial in that it makes a severe demand on the scientist and puts a severe limitation on the results. It says roughly this: If you can assign a degree of certainty, or personal probability, to the premises of your valid argument, you may use any and all the rules of probability theory to derive a certainty for the conclusion, and this certainty will be a logically valid consequence of your original certainties. The catch is that your concluding certainty, or *posterior probability,* may heavily depend on what you used as initial certainties, or *prior probabilities.* And if those initial certainties are not those of a colleague, that colleague may very well assign a certainty to the conclusion different from the one you derived.

Because the posterior probabilities emanating from a Bayesian inference depend on the person supplying the initial certainties and so may vary across individuals, the inferences are said to be subjective. This subjectivity of Bayesian inference is often mistaken for a subjective treatment of truth. Not only is such a view of Bayesianism incorrect, but it is diametrically opposed to Bayesian philosophy. The Bayesian approach represents a constructive attempt to deal with the dilemma that scientific laws and facts should not be treated as known with certainty, whereas classic deductive logic yields conclusions only when some law, fact, or connection is asserted with 100% certainty.

A common criticism of Bayesian philosophy is that it diverts attention away from the classic goals of science, such as the discovery of how the world works, toward psychologic states of mind called "certainties," "subjective probabilities," or "degrees of belief" (Popper, 1959). This criticism fails, however, to recognize the importance of a scientist's state of mind in determining what theories to test and what tests to apply.

In any research context, there will be an unlimited number of hypotheses that could explain an observed phenomenon. Some argue that progress is best aided by severely testing (empirically challenging) those explanations that seem most probable in light of past research, so that shortcomings of currently "received" theories can be most rapidly discovered. Indeed, much research in certain fields takes this form, as when theoretical predictions of particle mass are put to ever more precise tests in physics experiments. This process does not involve mere improved repetition of past studies. Rather, it involves tests of previously untested but important predictions of the theory.

Probabilities of auxiliary hypotheses are also important in study design and interpretation. Failure of a theory to pass a test can lead to rejection of the theory more rapidly when the auxiliary hypotheses on which the test depends possess high probability. This observation provides a rationale for preferring population-based to hospital-based case-control studies, because the former may have a higher probability of unbiased subject selection.

PM3003572236

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

Even if one disputes the above arguments, most epidemiologists desire some interval estimate or evaluation of the likely range for an effect in light of available data. This estimate must inevitably be derived in the face of considerable uncertainty about methodologic details and various events that led to the available data, and can be extremely sensitive to the reasoning used in its derivation. Psychologic investigations have found that most people, including scientists, reason poorly in general and especially poorly in the face of uncertainty (Kahneman et al., 1982, Piattelli-Palmarini, 1994). Bayesian philosophy provides a methodology for sound reasoning and, in particular, provides many warnings against being overly certain about one's conclusions (Greenland 1998ab).

Such warnings are echoed in refutationist philosophy. As Peter Medawar (1979) put it:

> I cannot give any scientist of any age better advice than this: the intensity of the conviction that a hypothesis is true has no bearing on whether it is true or not.

We would only add that intensity of conviction that a hypothesis is false has no bearing on whether it is false or not.

Impossibility of Scientific Proof

Vigorous debate is a characteristic of modern scientific philosophy, no less in epidemiology than in other areas (Rothman, 1988). Perhaps the most important common thread that emerges from the debated philosophies is Hume's legacy that proof is impossible in empiric science. This simple fact is especially important to epidemiologists, who often face the criticism that proof is impossible in epidemiology, with the implication that it is possible in other scientific disciplines. Such criticism may stem from a view that experiments are the definitive source of scientific knowledge. Such a view is mistaken on at least two counts. First, the nonexperimental nature of a science does not preclude impressive scientific discoveries; the myriad examples include plate tectonics, the evolution of species, planets orbiting other stars, and the effects of cigarette smoking on human health. Even when they are possible, experiments (including randomized trials) do not provide anything approaching proof and in fact may be controversial, contradictory, or irreproducible. The cold-fusion debacle demonstrates well that neither physical nor experimental science is immune to such problems (Taubes, 1993).

Some experimental scientists hold that epidemiologic relations are only suggestive and believe that detailed laboratory study of mechanisms within single individuals can reveal cause–effect relations with certainty. This view overlooks the fact that *all* relations are suggestive in exactly the manner discussed by Hume: Even the most careful and detailed mechanistic dissection of individual events cannot provide more than associations, albeit at a finer level. Laboratory studies often involve a degree of observer control that cannot be approached in epidemiology; it is only this control, not the level of observation, that can strengthen the inferences from laboratory studies. And again, such control is no guarantee against error.

All of the fruits of scientific work, in epidemiology or other disciplines, are at best only tentative formulations of a description of nature, even when the work itself is carried out without mistakes. The tentativeness of our knowledge does not prevent practical applications, but it should keep us skeptical and critical, not only of everyone else's work but of our own as well.

PM3003572237

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

Biologic knowledge about epidemiologic hypotheses is often scant, making the hypotheses themselves at times little more than vague statements of causal association between exposure and disease, such as "smoking causes cardiovascular disease." These vague hypotheses have only vague consequences that can be difficult to test. To cope with this vagueness, epidemiologists usually focus on testing the negation of the causal hypothesis, that is, the null hypothesis that the exposure does *not* have a causal relation to disease. Then, any observed association can potentially refute the hypothesis, subject to the assumption (auxiliary hypothesis) that biases are absent.

Testing Competing Epidemiologic Theories

If the causal mechanism is stated specifically enough, epidemiologic observations can provide crucial tests of competing non-null causal hypotheses. For example, when toxic shock syndrome was first studied, there were two competing hypotheses about the origin of the toxin. Under one hypothesis, the toxin was a chemical in the tampon, so that women using tampons were exposed to the toxin directly from the tampon. Under the other hypothesis, the tampon acted as a culture medium for staphylococci that produced the toxin. Both hypotheses explained the relation of toxic shock occurrence to tampon use. The two hypotheses, however, lead to opposite predictions about the relation between the frequency of changing tampons and the risk of toxic shock. Under the hypothesis of a chemical intoxication, more frequent changing of the tampon would lead to more exposure to the toxin and possible absorption of a greater overall dose. This hypothesis predicted that women who changed tampons more frequently would have a higher risk than women who changed tampons infrequently. The culture-medium hypothesis predicts that the women who change tampons frequently would have a lower risk than those who leave the tampon in for longer periods, because a short duration of use for each tampon would prevent the staphylococci from multiplying enough to produce a damaging dose of toxin. Thus, epidemiologic research, by showing that infrequent changing of tampons was associated with the risk of toxic shock, refuted the chemical theory.

Another example of a theory easily tested by epidemiologic data relates to the finding that women who took replacement estrogen therapy were at a considerably higher risk for endometrial cancer. Horwitz and Feinstein (1978) conjectured a competing theory to explain the association: They proposed that women taking estrogen experienced symptoms such as bleeding that induced them to consult a physician. The resulting diagnostic workup led to the detection of endometrial cancer at an earlier stage in these women, as compared with women not taking estrogens. Many epidemiologic observations could have been and were used to evaluate these competing hypotheses. The causal theory predicted that the risk of endometrial cancer would tend to increase with increasing use (dose, frequency, and duration) of estrogens, as for other carcinogenic exposures. The detection bias theory, on the other hand, predicted that women who had used estrogens only for a short while would have the greatest risk, since the symptoms related to estrogen use that led to the medical consultation tended to appear soon after use began. Because the association of recent estrogen use and endometrial cancer was the same in both long- and short-term estrogen users, the detection bias theory was refuted as an explanation for all but a small fraction of endometrial cancer cases occurring after estrogen use. (Refutation of the detection bias theory also depended on many other observations. Especially im-

PM3003572238

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

portant was the theory's implication that there must be a large reservoir of undetected endometrial cancer in the typical population of women to account for the much greater rate observed in estrogen users.)

The endometrial cancer example illustrates a critical point in understanding the process of causal inference in epidemiologic studies: Many of the hypotheses being evaluated in the interpretation of epidemiologic studies are auxiliary hypotheses in the sense of involving no causal connection between the study exposure and the disease. For example, hypotheses that amount to explanations of how specific types of bias could have led to an association between exposure and disease are the usual alternatives to the primary study hypothesis that the epidemiologist needs to consider in drawing inferences. Much of the interpretation of epidemiologic studies amounts to the testing of such auxiliary explanations for observed associations.

Causal Criteria

In practice, how do epidemiologists separate causal from noncausal explanations? Despite philosophic criticisms of inductive inference, inductively oriented causal criteria are often used to make such inferences. If a set of necessary and sufficient causal criteria could be used to distinguish causal from noncausal relations in epidemiologic studies, the job of the scientist would be eased considerably. With such criteria, all the concerns about the logic or lack thereof in causal inference could be forgotten: It would only be necessary to consult the checklist of criteria to see if a relation were causal. We know from philosophy that a set of sufficient criteria does not exist. Nevertheless, lists of causal criteria have become popular, possibly because they seem to provide a road map through complicated territory.

A commonly used set of criteria was proposed by Hill (1965); it was an expansion of a set of criteria offered previously in the landmark U.S. Surgeon General's report *Smoking and Health* (1964), which in turn was anticipated by the inductive canons of John Stuart Mill (1862) and the rules given by Hume (1739). Hill suggested that the following aspects of an association be considered in attempting to distinguish causal from noncausal associations: (1) strength, (2) consistency, (3) specificity, (4) temporality, (5) biologic gradient, (6) plausibility, (7) coherence, (8) experimental evidence, and (9) analogy. The popular view that these criteria should be used for causal inference makes it necessary to examine them in detail:

1. **Strength.** Hill argued that strong associations are more likely to be causal than weak associations because, if they could be explained by some other factor, the effect of that factor would have to be even stronger than the observed association and therefore would have become evident. Weak associations, on the other hand, are more likely to be explained by undetected biases. To some extent, this is a reasonable argument, but, as Hill himself acknowledged, the fact that an association is weak does not rule out a causal connection. A commonly cited counterexample is the relation between cigarette smoking and cardiovascular disease: One explanation for this relation being weak is that cardiovascular disease is common, making any ratio measure of effect comparatively small compared with ratio measures for diseases that are less common (Rothman and Poole, 1988). Nevertheless, cigarette smoking is not seriously doubted as a cause of cardiovascular disease. Another example would be passive smoking and lung cancer, a weak association that few consider to be noncausal.

PM3003572239

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

Counterexamples of strong but noncausal associations are also not hard to find; any study with strong confounding illustrates the phenomenon. For example, consider the strong but noncausal relation between Down syndrome and birth rank, which is confounded by the relation between Down syndrome and maternal age. Of course, once the confounding factor is identified, the association is diminished by adjustment for the factor. These examples remind us that a strong association is neither necessary nor sufficient for causality, and that weakness is neither necessary nor sufficient for absence of causality. In addition to these counterexamples, we have to remember that neither relative risk nor any other measure of association is a biologically consistent feature of an association; as described earlier in this chapter, it is a characteristic of a study population that depends on the relative prevalence of other causes. A strong association serves only to rule out hypotheses that the association is entirely due to one weak unmeasured confounder or other source of modest bias.

2. **Consistency.** Consistency refers to the repeated observation of an association in different populations under different circumstances. Lack of consistency, however, does not rule out a causal association because some effects are produced by their causes only under unusual circumstances. More precisely, the effect of a causal agent cannot occur unless the complementary component causes act or have already acted to complete a sufficient cause. These conditions will not always be met. Thus, transfusions can cause infection with the human immunodeficiency virus, but they do not always do so: The virus must also be present. Tampon use can cause toxic shock syndrome, but only rarely, when certain other, perhaps unknown, conditions are met. Consistency is apparent only after all the relevant details of a causal mechanism are understood, which is to say very seldom. Furthermore, even studies of exactly the same phenomena can be expected to yield different results simply because they differ in their methods and random errors. Consistency serves only to rule out hypotheses that the association is attributable to some factor that varies across studies.
3. **Specificity.** The criterion of specificity requires that a cause lead to a single effect, not multiple effects. This argument has often been advanced to refute causal interpretations of exposures that appear to relate to myriad effects, especially by those seeking to exonerate smoking as a cause of lung cancer. Unfortunately, the criterion is wholly invalid. Causes of a given effect cannot be expected to lack other effects on any logical grounds. In fact, everyday experience teaches us repeatedly that single events or conditions may have many effects. Smoking is an excellent example; it leads to many effects in the smoker. The existence of one effect does not detract from the possibility that another effect exists.

   To summarize, specificity does not confer greater validity to any causal inference regarding the exposure effect. Hill's discussion of this criterion for inference is replete with reservations, but even so, the criterion is useless and misleading.
4. **Temporality.** Temporality refers to the necessity that the cause precede the effect in time. This criterion is inarguable, insofar as any claimed observation of causation must involve the putative cause C preceding the putative effect D. It does *not*, however, follow that a reverse time order is evidence against the hypothesis that C can cause D. Rather, observations in which C followed D merely show that C could not have caused D in these instances; they provide no evidence for or against the hypothesis that C can cause D in those instances in which it precedes D.
5. **Biologic gradient.** Biologic gradient refers to the presence of a monotonic (unidirectional) dose–response curve. We often expect such a monotonic relation to exist. For example, more smoking means more carcinogen exposure and more tissue dam-

PM3003572240

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

age, hence more opportunity for carcinogenesis. Some causal associations, however, show a single jump (threshold) rather than a monotonic trend; an example is the association between DES and adenocarcinoma of the vagina. A possible explanation is that the doses of DES that were administered were all sufficiently great to produce the maximum effect from DES. Under this hypothesis, for all those exposed to DES, the development of disease would depend entirely on other component causes.

The somewhat controversial topic of alcohol consumption and mortality is another example. Death rates are higher among nondrinkers than among moderate drinkers, but they ascend to the highest levels for heavy drinkers. There is considerable debate about which parts of the J-shaped dose–response curve are causally related to alcohol consumption and which parts are noncausal artifacts stemming from confounding or other biases. Some studies appear to find only an increasing relation between alcohol consumption and mortality, possibly because the categories of alcohol consumption are too broad to distinguish different rates among moderate drinkers and nondrinkers.

Associations that do show a monotonic trend in disease frequency with increasing levels of exposure are not necessarily causal; confounding can result in a monotonic relation between a noncausal risk factor and disease if the confounding factor itself demonstrates a biologic gradient in its relation with disease. The noncausal relation between birth rank and Down syndrome mentioned above shows a biologic gradient that merely reflects the progressive relation between maternal age and Down syndrome occurrence.

These issues imply that the existence of a monotonic association is neither necessary nor sufficient for a causal relation. A nonmonotonic relation only refutes those causal hypotheses specific enough to predict a monotonic dose–response curve.

6. **Plausibility.** Plausibility refers to the biologic plausibility of the hypothesis, an important concern but one that is far from objective or absolute. Sartwell (1960) emphasized this point, citing the remarks of Cheever in 1861, who had been commenting on the etiology of typhus before its mode of transmission (via body lice) was even known:

> It could be no more ridiculous for the stranger who passed the night in the steerage of an emigrant ship to ascribe the typhus, which he there contracted, to the vermin with which bodies of the sick might be infested. An adequate cause, one reasonable in itself, must correct the coincidences of simple experience.

What was to Cheever an implausible explanation turned out to be the correct explanation, since it was indeed the vermin that caused the typhus infection. Such is the problem with plausibility: It is too often not based on logic or data, but only on prior beliefs. This is not to say that biologic knowledge should be discounted when a new hypothesis is being evaluated, but only to point out the difficulty in applying that knowledge.

The Bayesian approach to inference attempts to deal with this problem by requiring that one quantify, on a probability (zero to one) scale, the certainty that one has in prior beliefs, as well as in new hypotheses. This quantification displays the dogmatism or open-mindedness of the analyst in a public fashion, with certainty values near one or zero betraying a strong commitment of the analyst for or against a hypothesis. It can also provide a means of testing those quantified beliefs against new evidence (Howson and Urbach, 1993). Nevertheless, the Bayesian approach cannot transform plausibility into an objective causal criterion.

PM3003572241

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

7. **Coherence.** Taken from the U.S. Surgeon General's *Smoking and Health* (1964), the term *coherence* implies that a cause-and-effect interpretation for an association does not conflict with what is known of the natural history and biology of the disease. The examples Hill gave for coherence, such as the histopathologic effect of smoking on bronchial epithelium (in reference to the association between smoking and lung cancer) or the difference in lung cancer incidence by sex, could reasonably be considered examples of plausibility, as well as coherence; the distinction appears to be a fine one. Hill emphasized that the absence of coherent information, as distinguished, apparently, from the presence of conflicting information, should not be taken as evidence against an association being considered causal. On the other hand, presence of conflicting information may indeed refute a hypothesis, but one must always remember that the conflicting information may be mistaken or misinterpreted (Wald, 1985).
8. **Experimental evidence.** It is not clear what Hill meant by experimental evidence. It might have referred to evidence from laboratory experiments on animals or to evidence from human experiments. Evidence from human experiments, however, is seldom available for most epidemiologic research questions, and animal evidence relates to different species and usually to levels of exposure very different from those humans experience. From Hill's examples, it seems that what he had in mind for experimental evidence was the result of removal of some harmful exposure in an intervention or prevention program, rather than the results of laboratory experiments (Susser, 1991). The lack of such evidence would at least be a pragmatic difficulty in making this a criterion for inference. Logically, however, experimental evidence is not a criterion, but a test of the causal hypothesis, a test that is simply unavailable in most circumstances. Although experimental tests can be much stronger than other tests, they are not as decisive as often thought because of difficulties in interpretation. For example, one can attempt to test the hypothesis that malaria is caused by swamp gas by draining swamps in some areas and not in others to see if the malaria rates among residents are affected by the draining. As predicted by the hypothesis, the rates will drop in the areas where the swamps are drained. As Popper emphasized, however, there are always many alternative explanations for the outcome of every experiment. In this example, one alternative, which happens to be correct, is that mosquitoes are responsible for malaria transmission.
9. **Analogy.** Whatever insight might be derived from analogy is handicapped by the inventive imagination of scientists who can find analogies everywhere. At best, analogy provides a source of more elaborate hypotheses about the associations under study; absence of such analogies only reflects lack of imagination or experience, not falsity of the hypothesis.

As is evident, the standards of epidemiologic evidence offered by Hill are saddled with reservations and exceptions. Hill himself was ambivalent about the utility of these "standards" (he did not use the word *criteria* in the paper). On the one hand, he asked, "In what circumstances can we pass from this observed *association* to a verdict of *causation*?" (original emphasis). Yet, despite speaking of verdicts on causation, he disagreed that any "hard-and-fast rules of evidence" existed by which to judge causation:

> None of my nine viewpoints [criteria] can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*.

Actually, the fourth criterion, temporality, is a *sine qua non* for causality: If the putative cause did not precede the effect, that indeed is indisputable evidence that the ob-

PM3003572242

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001

served association is not causal (although this evidence does not rule out causality in other situations, for in other situations the putative cause may precede the effect). Other than this one condition, however, which may be viewed as part of the definition of causation, there is no necessary or sufficient criterion for determining whether an observed association is causal.

This conclusion accords with the views of Hume, Popper, and others that causal inferences cannot attain the certainty of logical deductions. Although some scientists continue to promulgate causal criteria as aids to inference (Susser, 1991), others argue that it is actually detrimental to cloud the inferential process by considering checklist criteria (Lanes and Poole, 1984). An intermediate, refutationist approach seeks to transform the criteria into deductive tests of causal hypotheses (Maclure, 1985; Weed, 1986). Such an approach avoids the temptation to use causal criteria simply to buttress pet theories at hand, and instead allows epidemiologists to focus on evaluating competing causal theories using crucial observations.

PM3003572243

Source: https://www.industrydocuments.ucsf.edu/docs/msmj0001